TITLE OF THE INVENTION

Method And Polynucleotides For Determining Translational Efficiency Of A Codon

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application of co-pending International Patent Application No. PCT/AU00/00008 filed January 7, 2000, which designates the United States, and which claims priority of Australian Provisional Patent Application No. PP8078/99 filed January 8, 1999.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] Not applicable.

REFERENCE TO A MICROFICHE APPENDIX

[0003] Not applicable.

BACKGROUND OF THE INVENTION

[0004] This invention relates generally to gene expression and in particular, to a method and polynucleotides for determining codon utilization in particular cells or tissues of an organism. More particularly, the method and polynucleotides of the invention are concerned with ascertaining codon preferences in cells or tissues for the purpose of modifying the translational efficiency of protein-encoding polynucleotides in those cells or tissues.

[0005] It is well known that a "triplet" codon of four possible nucleotide bases can exist in 64 variant forms. These forms provide the message for only 20 different amino acids (as well as translation initiation and termination) and this means

that some amino acids can be encoded by more than one codon. Some amino acids have as many as six "redundant", alternative codons while some others have a single, required codon.

[0006] For reasons not completely understood, codon utilization is highly biased in that alternative codons are not at all uniformly present in the endogenous DNA of differing cell types. In this regard, there appears to exist a variable natural hierarchy of "preference" for certain codons between different cell types or between different organisms.

[0007] Codon usage patterns have been shown to correlate with relative abundance of isoaccepting transfer RNA (iso-tRNA) species, and with genes encoding proteins of high versus low abundance. Moreover, the present inventors recently discovered that the intracellular abundance of different iso-tRNAs varies in different cells or tissues of a single multi-cellular organism (see copending International Application No. PCT/AU98/00530).

[0008] The implications of codon preference phenomena on gene expression are manifest in that these phenomena can affect the translational efficiency of messenger RNA (mRNA). It is widely known in this regard that translation of "rare codons", for which the corresponding iso-tRNA is in relatively low abundance, may cause a ribosome to pause during translation which can lead to a failure to complete a nascent polypeptide chain and an uncoupling of transcription and translation.

[0009] A primary goal in recombinant research is to provide transgenic organisms with expression of a foreign gene in an amount sufficient to confer the desired phenotype to the organism. However, expression of the foreign gene may be severely impeded if a particular host cell of the organism or the organism itself has a low abundance of iso-tRNAs corresponding to one or more codons of the foreign gene. Accordingly, a major aim of investigators in this field is to

first ascertain the codon preference for particular cells or tissues in which a foreign gene is to be expressed, and to subsequently alter the codon composition of the foreign gene for optimized expression in those cells or tissues.

[0010] Codon preference may be determined simply by analyzing the frequency at which codons are used by genes expressed in a particular cell or tissue or in a plurality of cells or tissues of a given organism. Codon frequency tables as well as suitable methods for determining frequency of codon usage in an organism are described, for example, in an article by Sharp *et al* (1988, *Nucleic Acids Res.* 16 8207-8211). The relative level of gene expression (*e.g.*, detectable protein expression Vs no detectable protein expression) can provide an indirect measure of the relative abundance of specific iso-tRNAs expressed in different cells or tissues.

[0011] Alternatively, codon preference may be determined by measuring the relative intracellular abundance of different iso-tRNA species. For example, reference may be made to copending International Application No. PCT/AU98/00530 that describes a method that utilizes labeled oligonucleotides specific for different iso-tRNAs to probe an RNA extract prepared from a particular cell or tissue source.

[0012] The above methods provide useful indirect evidence for determining codon preference. However, such indirect evidence may not provide an accurate indication of the translational efficiency of a given codon. Accordingly, there is a need to provide a method that more directly ascertains the translational efficiency of a codon in a cell or tissue.

SUMMARY OF THE INVENTION

[0013] In one aspect of the invention, there is provided a method for determining the translational efficiency of an

individual codon in a cell of a predetermined type, said method comprising:

- introducing into a first cell of said predetermined type a synthetic construct comprising a reporter polynucleotide fused in frame with a tandem repeat of said individual codon, wherein said reporter polynucleotide encodes a reporter protein, and wherein said synthetic construct is operably linked to a regulatory polynucleotide; and

- measuring expression of said reporter protein in said cell of said predetermined type to determine the translational efficiency of said codon.

[0014] Preferably, the method further comprises comparing:

- expression of said reporter protein in said first cell to which a synthetic construct comprising a tandem repeat of said individual codon was provided; and

- expression of said reporter protein in second cell of the same type as said first cell to which a synthetic construct comprising a tandem repeat of another individual codon was provided;

to thereby determine the relative translational efficiency of said individual codons in said cell of said predetermined type.

[0015] Suitably, the method further comprises comparing:

- expression of said reporter protein in said first cell to which a synthetic construct comprising a tandem repeat of said individual codon was provided; and

- expression of said reporter protein in another cell of a different predetermined type than said first cell to which a synthetic construct comprising a tandem repeat of said individual codon was provided;

to thereby determine the translational efficiency of said individual codon in said first cell relative to said other cell.

[0016] Preferably, the method further comprises:

- introducing the synthetic construct into a progenitor cell of said cell of said predetermined type; and

- producing said cell of said predetermined type from said progenitor cell;

wherein said cell of said predetermined type contains said synthetic construct.

[0017] Suitably, the method further comprises:

- introducing the synthetic construct into a progenitor of said cell; and

- growing an organism or part thereof from said progenitor cell;

wherein said organism or part thereof comprises said cell containing said synthetic construct.

[0018] Suitably, the method further comprises:

- introducing the synthetic construct into an organism or part thereof such that said synthetic construct is introduced into said cell of said predetermined type.

[0019] In another aspect, the invention resides in a synthetic construct comprising a reporter polynucleotide fused in frame with a tandem repeat of individual codons, wherein said reporter polynucleotide encodes a reporter protein, and wherein said synthetic construct is operably linked to a regulatory polynucleotide.

[0020] In yet another aspect of the invention, there is provided a method of constructing a synthetic polynucleotide from which a protein is selectively expressed in a target cell of an organism, relative to another cell of the organism, said method comprising:

- selecting a first codon of a parent polynucleotide for replacement with a synonymous codon which has a higher translational efficiency in said target cell than in said other cell; and

- replacing said first codon with said synonymous codon to form said synthetic polynucleotide, wherein said first codon and said synonymous codon are selected by:

  - comparing translational efficiencies of individual codons in said target cell relative to said other cell using the method broadly described above; and

  - selecting said first codon and said synonymous codon based on said comparison.

[0021] Preferably, said synonymous codon corresponds to a reporter construct from which the reporter protein is expressed in said target cell at a level that is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from said reporter construct in said other cell.

[0022] In a further aspect, the invention provides a method of constructing a synthetic polynucleotide from which a protein is expressible in a target cell of an organism at a higher level than from a parent polynucleotide encoding said protein, said method comprising:

- selecting a first codon of the parent polynucleotide for replacement with a synonymous codon which has a higher translational efficiency in said target cell than said first codon;

- replacing said first codon with said synonymous codon to form said synthetic polynucleotide, wherein said first codon and said synonymous codon are selected by:

- comparing translational efficiencies of different individual codons in said target cell using the method broadly described above; and

- selecting said first codon and said synonymous codon based on said comparison.

[0023] Suitably, said synonymous codon corresponds to a reporter construct from which the reporter protein is expressed in said target cell at a level that is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from the different reporter construct corresponding to said first codon.

DETAILED DESCRIPTION OF THE INVENTION

1. Definitions

[0024] The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0025] Throughout this specification, unless the context requires otherwise, the words "comprise", "comprises" and "comprising" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

[0026] By "expressible" is meant expression of a protein to a level sufficient to effect a particular function associated with the protein. By contrast, the terms "not expressible" and "not substantially expressible" as used interchangeably herein refers to (a) no expression of a protein, (b) expression of a protein to a level that is not sufficient to effect a particular function associated with the protein, (c) expression of a protein, which cannot be detected by a monoclonal antibody

specific for the protein, or (d) expression of a protein, which is less that 1% of the level expressed in a wild-type cell that normally expresses the protein.

[0027] By "expressing said synthetic construct" is meant transcribing the synthetic construct such that mRNA is produced.

[0028] By "expression vector" is meant any autonomous genetic element capable of directing the synthesis of a protein encoded by the vector. Such expression vectors are known by practitioners in the art.

[0029] As used herein, the term "function" refers to a biological, enzymatic, or therapeutic function.

[0030] By "highly expressed genes" is meant genes that express high levels of mRNA, and preferably high level of protein, relative to other genes.

[0031] By "isoaccepting transfer RNA" or "iso-tRNA" is meant one or more transfer RNA molecules that differ in their anticodon nucleotide sequence but are specific for the same amino acid.

[0032] By "natural gene" is meant a gene that naturally encodes the protein. However, it is possible that the parent polynucleotide encodes a protein that is not naturally-occurring but has been engineered using recombinant techniques.

[0033] The term "non-cycling cell" as used herein refers to a cell that has withdrawn from the cell cycle and has entered the G0 state. In this state, it is known that transcription of endogenous genes and protein translation are at substantially reduced levels compared to phases of the cell cycle, namely G1, S, G2 and M. By contrast, the term "cycling cell" as used herein refers to a cell, which is in one of the above phases of the cell cycle.

[0034] By "obtained from" is meant that a sample such as, for example, a polynucleotide extract or polypeptide extract is

isolated from, or derived from, a particular source of the host. For example, the extract can be obtained from a tissue or a biological fluid isolated directly from the host.

[0035] The term "oligonucleotide" as used herein refers to a polymer composed of a multiplicity of nucleotide residues (deoxyribonucleotides or ribonucleotides, or related structural variants or synthetic analogues thereof) linked via phosphodiester bonds (or related structural variants or synthetic analogues thereof). Thus, while the term "oligonucleotide" typically refers to a nucleotide polymer in which the nucleotide residues and linkages between them are naturally occurring, it will be understood that the term also includes within its scope various analogues including, but not restricted to, peptide nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. The exact size of the molecule can vary depending on the particular application. An oligonucleotide is typically rather short in length, generally from about 10 to 30 nucleotide residues, but the term can refer to molecules of any length, although the term "polynucleotide" or "nucleic acid" is typically used for large oligonucleotides.

[0036] By "operably linked" is meant that transcriptional and translational regulatory polynucleotides are positioned relative to a polypeptide-encoding polynucleotide in such a manner that the polynucleotide is transcribed and the polypeptide is translated.

[0037] By "pharmaceutically-acceptable carrier" is meant a solid or liquid filler, diluent or encapsulating substance that can be safely used in topical or systemic administration to a mammal.

[0038] "Polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same.

Thus, these terms apply to amino acid polymers in which one or more amino acid residues is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers.

[0039] The term "polynucleotide" or "nucleic acid" as used herein designates mRNA, RNA, cRNA, cDNA or DNA. The term typically refers to oligonucleotides greater than 30 nucleotide residues in length.

[0040] By "primer" is meant an oligonucleotide which, when paired with a strand of DNA, is capable of initiating the synthesis of a primer extension product in the presence of a suitable polymerizing agent. The primer is preferably single-stranded for maximum efficiency in amplification but can alternatively be double-stranded. A primer must be sufficiently long to prime the synthesis of extension products in the presence of the polymerization agent. The length of the primer depends on many factors, including application, temperature to be employed, template reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 or more nucleotide residues, although it can contain fewer nucleotide residues. Primers can be large polynucleotides, such as from about 200 nucleotide residues to several kilobases or more. Primers can be selected to be "substantially complementary" to the sequence on the template to which it is designed to hybridize and serve as a site for the initiation of synthesis. By "substantially complementary", it is meant that the primer is sufficiently complementary to hybridize with a target polynucleotide. Preferably, the primer contains no mismatches with the template to which it is designed to hybridize but this is not essential. For example, non-complementary nucleotide residues can be attached to the 5′ end of the primer, with the remainder of the primer sequence being

complementary to the template. Alternatively, non-complementary nucleotide residues or a stretch of non-complementary nucleotide residues can be interspersed into a primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize therewith and thereby form a template for synthesis of the extension product of the primer.

[0041] "Probe" refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another molecule. Unless otherwise indicated, the term "probe" typically refers to a polynucleotide probe that binds to another polynucleotide, often called the "target polynucleotide", through complementary base pairing. Probes can bind target polynucleotides lacking complete sequence complementarity with the probe, depending on the stringency of the hybridization conditions. Probes can be labeled directly or indirectly.

[0042] The terms "precursor cell or tissue" and "progenitor cell or tissue" as used herein refer to a cell or tissue that can gives rise to a particular cell or tissue in which protein expression is to be targeted or in which translational efficiency of a codon is to be determined.

[0043] By "recombinant polypeptide" is meant a polypeptide made using recombinant techniques, i.e., through the expression of a recombinant or synthetic polynucleotide.

[0044] "Stringency" as used herein, refers to the temperature and ionic strength conditions, and presence or absence of certain organic solvents, during hybridization. The higher the stringency, the higher will be the degree of complementarity between immobilized polynucleotides and the labeled polynucleotide.

[0045] "Stringent conditions" refers to temperature and ionic conditions under which only polynucleotides having a high frequency of complementary bases will hybridize. The stringency required is nucleotide sequence dependent and depends upon the

various components present during hybridization. Generally, stringent conditions are selected to be about 10 to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a target sequence hybridizes to a complementary probe.

[0046] The term "synthetic polynucleotide" as used herein refers to a polynucleotide formed in vitro by the manipulation of a polynucleotide into a form not normally found in nature. For example, the synthetic polynucleotide can be in the form of an expression vector. Generally, such expression vectors include transcriptional and translational regulatory polynucleotide operably linked to the polynucleotide.

[0047] The term "synonymous codon" as used herein refers to a codon having a different nucleotide sequence than another codon but encoding the same amino acid as that other codon.

[0048] By "translational efficiency" is meant the efficiency of a cell's protein synthesis machinery to incorporate the amino acid encoded by a codon into a nascent polypeptide chain. This efficiency can be evidenced, for example, by the rate at which the cell is able to synthesize the polypeptide from an RNA template comprising the codon, or by the amount of the polypeptide synthesized from such a template.

[0049] By "vector" is meant a polynucleotide molecule, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, or plant virus, into which a polynucleotide can be inserted or cloned. A vector preferably contains one or more unique restriction sites and can be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector can be an autonomously replicating vector, i.e., a vector that exists as

an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a linear or closed circular plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector can contain any means for assuring self-replication. Alternatively, the vector can be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. A vector system can comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector can also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are known to those of skill in the art and include the nptII gene that confers resistance to the antibiotics kanamycin and G418 (Geneticin®) and the hph gene which confers resistance to the antibiotic hygromycin B.

2. Method of the invention

[0050] The present invention is based, at least in part, on the discovery that different but synonymous stretches of identical codons fused respectively in frame with a reporter polynucleotide can give rise to different levels of reporter protein expressed within a given cell type. Not wishing to be bound by any particular theory, it is believed that a tandem series of identical codons causes a ribosome to pause during translation if the iso-tRNA corresponding to the identical codons is limiting. In this regard, it is known that ribosomal pausing leads to a failure to complete a nascent polypeptide chain and an uncoupling of transcription and translation. Accordingly, the levels of reporter protein expressed in the

different cells or tissues are sensitive to the intracellular abundance of the iso-tRNA species corresponding to the identical codons and, therefore, provide a direct correlation of a cell's or tissue's preference for translating a given codon. This means, for example, that if the levels of the reporter protein obtained in a cell or tissue type to which a synthetic construct having a tandem series of identical first codons is provided are lower than the levels expressed in the same cell or tissue type to which a different synthetic construct having a tandem series of identical second codons is provided (i.e., wherein the first codons are different from, but synonymous with, the second codons), then it can be deduced that the cell or tissue has a higher preference for the second codon relative to the first codon with respect to translation. Put another way, the second codon has a higher translational efficiency compared to the first codon in the cell or tissue type.

[0051] With regard to differential protein expression between different cell or tissue types, it will be appreciated that if the levels of the reporter protein obtained in a target cell or tissue type to which a synthetic construct having a tandem series of identical codons is provided are lower than the levels expressed in the another cell or tissue type to which the same synthetic construct is provided, then it can be deduced that the target cell or tissue has a higher preference for the codon relative to the other cell or tissue with respect to translation. Put another way, the codon has a higher translational efficiency in the target cell or tissue relative to the other cell or tissue type.

[0052] As used herein, expression of a protein in a tissue refers alternatively to expression of the protein within a cell of the tissue or production of the protein within a cell and export of the protein from the cell to, for example, the extracellular matrix of a tissue.

[0053] Suitably, the tandem repeat comprises at least three identical codons. Preferably, the tandem repeat comprises four identical codons, more preferably five or seven identical codons and most preferably six identical codons.

[0054] The tandem repeat can be fused at a location adjacent to, or within, the reporter polynucleotide. The location is preferably selected such that the tandem repeat interferes with translation of at least a detectable portion of the reporter protein such that expression of the protein can be detected or assessed. Preferably, the tandem repeat is located immediately upstream (translationally) from the reporter polynucleotide.

[0055] It is of course possible that a tandem repeat of identical amino acid residues (e.g., an oligo-proline repeat) can render the reporter protein unstable. Typically, protein instability is detected when expression of the reporter gene is not detectable with any choice of isoaccepting codon specific for the amino acid corresponding to the tandem repeat. The inventors have found in this regard that protein instability can be alleviated by use of at least one spacer codon within the tandem repeat of identical codons, wherein the spacer codon encodes a neutral amino acid.

[0056] The at least one spacer codon can be placed adjacent to, or interposed between, some or all of the identical codons corresponding to the tandem repeat. For example, a suitable interposition for a penta-repeat of identical codons can be selected from the group consisting of: (a) I-S-I-S-I-S-I-S-I-S; (b) S-I-S-I-S-I-S-I-S-I; (c) I-S-I-S-I-I-S-I; (d) I-S-I-I-S-I-S-I; (e) I-S-I-S-I-I-I; (f) I-I-S-I-S-I-I; (g) I-I-I-S-I-S-I; (h) I-S-I-I-S-I-I; (i) I-I-S-I-I-S-I; (j) I-S-I-I-I-S-I; (k) I-S-I-I-I-I; (l) I-I-S-I-I-I; (m) I-I-I-S-I-I; and (n) I-I-I-I-S-I, wherein I corresponds to an identical codon of a tandem repeat and S corresponds to a spacer codon.

[0057] Preferably, a spacer codon is efficiently translated in the cell or tissue type relative to other synonymous codons. This is important so that translation of the spacer codon is not rate limiting. The neutral amino acid includes, but is not restricted to, alanine and glycine.

[0058] The reporter polynucleotide can encode any suitable protein for which expression can be detected directly or indirectly such as by suitable assay. Suitable reporter polynucleotides include, but are not restricted to, polynucleotides encoding β-galactosidase, firefly luciferase, alkaline phosphatase, chloramphenicol acetyltransferase (CAT), β-glucuronidase (GUS), herbicide resistance genes such as the bialophos resistance (BAR) gene that confers resistance to the herbicide BASTA, and green fluorescent protein (GFP). Assays for the activities associated with such proteins are known by those of skill in the art. Preferably, the reporter polynucleotide encodes GFP.

[0059] Persons of skill in the art will appreciate that reporter polynucleotides need not correspond to a full-length gene encoding a particular reporter protein. In this regard, the invention also contemplates reporter polynucleotide sub-sequences encoding desired portions of a parent reporter protein, wherein an activity or function of the parent protein is retained in said portions. A polynucleotide sub-sequence encodes a domain of the reporter protein having an activity associated therewith and preferably encodes at least 10, 20, 50, 100, 150, or 500 contiguous amino acid residues of the reporter protein.

[0060] The instant method is applicable to any suitable cell or tissue type and, hence, is not restricted to application to mammalian cells/tissues. Accordingly, the cell or tissue type can be of any animal or plant origin. The cell or tissue type can be of any suitable lineage. For example, a suitable cell

can include a eukaryotic cell, and preferably a cell or cell line capable of being grown in vitro. Suitable cell lines can include, for example, CV-1 cells, COS cells, yeast or spodoptera cells. The invention also contemplates cells that can be prokaryotic in origin.

[0061] Suitable methods for isolating particular cells or tissues are known to those of skill in the art. For example, one can take advantage of one or more particular characteristics of a cell or tissue to specifically isolate the cell or tissue from a heterogeneous population. Such characteristics include, but are not limited to, anatomical location of a tissue, cell density, cell size, cell morphology, cellular metabolic activity, cell uptake of ions such as Ca^{2+}, K^{+}, and H^{+} ions, cell uptake of compounds such as stains, markers expressed on the cell surface, protein fluorescence, and membrane potential. Suitable methods that can be used in this regard include surgical removal of tissue, flow cytometry techniques such as fluorescence-activated cell sorting (FACS), immunoaffinity separation (e.g., magnetic bead separation such as Dynabead™ separation), density separation (e.g., metrizamide, Percoll™, or Ficoll™ gradient centrifugation), and cell-type specific density separation.

[0062] In an alternate embodiment, progenitor cells or tissues can be used for initially introducing the synthetic construct. Any suitable progenitor cell or tissue can be used which gives rise to a particular cell or tissue of interest for which codon preference is to be ascertained. For example, a suitable progenitor cell can comprise an undifferentiated cell. In the case of a plant, a suitable progenitor cell and tissue can include a meristematic cell and a callus tissue, respectively.

[0063] In another embodiment, the synthetic construct can be introduced first into an organism or part thereof before

subsequent expression of the construct in a particular cell or tissue type. Any suitable organism is contemplated by the invention including unicellular and as multi-cellular organisms. Exemplary multi-cellular organisms include plants and animals such as mammals(e.g., humans).

[0064] The invention further provides a synthetic construct comprising a reporter polynucleotide fused in frame with a tandem repeat of (e.g., 2, 3, 4, 5, 6, or 7 or more) identical codons, wherein said reporter polynucleotide encodes a reporter protein, and wherein said synthetic construct is operably linked to one or more regulatory polynucleotides.

[0065] The construction of the synthetic construct can be effected by any suitable technique. For example, in vitro mutagenesis methods can be employed, which are known to those of skill in the art. Suitable mutagenesis methods are described for example in the relevant sections of Ausubel, et al. (supra) and of Sambrook, et al., (supra) which are incorporated herein by reference. Alternatively, suitable methods for altering DNA are set forth, for example, in U.S. Patent Nos. 4,184,917, 4,321,365 and 4,351,901, which are incorporated herein by reference. Instead of in vitro mutagenesis, the synthetic construct can be synthesized de novo using readily available machinery. Sequential synthesis of DNA is described, for example, in U.S. Patent No 4,293,652, which is incorporated herein by reference. However, it should be noted that the present invention is not dependent on, and not directed to, any one particular technique for constructing the synthetic construct.

[0066] Regulatory polynucleotides which can be utilized to regulate expression of the synthetic construct include, but are not limited to, a promoter, an enhancer, and a transcriptional terminator. Such regulatory polynucleotides are known to those of skill in the art. The construct preferably comprises at

least one promoter. Suitable promoters that can be utilized to induce expression of the polynucleotides of the invention include constitutive promoters and inducible promoters.

[0067] The step of introducing the synthetic construct into a particular cell or tissue type, or into a progenitor cell or tissue thereof, or into an organism or part thereof for subsequent introduction into a particular cell or tissue will differ depending on the intended use and or species, and may involve lipofection, electroporation, micro-projectile bombardment infection by Agrobacterium tumefaciens or A rhizogenes, or protoplast fusion. Such methods are known to those skilled in the art.

[0068] Alternatively, the step of introduction may involve non-viral and viral vectors, cationic liposomes, retroviruses and adenoviruses such as, for example, described in Mulligan, R.C., (1993 Science 260 926-932) which is incorporated herein by reference. Such methods may include:

[0069] A. Local application of the synthetic nucleic acid sequence by injection (Wolff et al., 1990, Science 247 1465-1468, which is incorporated herein by reference), surgical implantation, instillation or any other means. This method may also be used in combination with local application by injection, surgical implantation, instillation or any other means, of cells responsive to the reporter protein encoded by the synthetic construct. This method may also be used in combination with local application by injection, surgical implantation, instillation or any other means, of another factor or factors required for the activity of said reporter protein.

[0070] B. General systemic delivery by injection of DNA, (Calabretta et al., 1993, Cancer Treat. Rev. 19 169-179, which is incorporated herein by reference), or RNA, alone or in combination with liposomes (Zhu et al., 1993, Science 261 209-212, which is incorporated herein by reference), viral capsids

or nanoparticles (Bertling et al., 1991, Biotech. Appl. Biochem. 13 390-405, which is incorporated herein by reference) or any other mediator of delivery. Improved targeting might be achieved by linking the synthetic construct to a targeting molecule (the so-called "magic bullet" approach employing for example, an antibody), or by local application by injection, surgical implantation or any other means, of another factor or factors required for the activity of the protein produced from said synthetic construct, or of cells responsive to said reporter protein.

[0071] C. Injection or implantation or delivery by any means, of cells that have been modified ex vivo by transfection (for example, in the presence of calcium phosphate: Chen et al., 1987, Mole. Cell Biochem. 7 2745-2752, or of cationic lipids and polyamines: Rose et al., 1991, BioTech. 10 520-525, which articles are incorporated herein by reference), infection, injection, electroporation (Shigekawa et al., 1988, BioTech. 6 742-751, which is incorporated herein by reference) or any other way so as to increase the expression of said synthetic construct in those cells. The modification may be mediated by plasmid, bacteriophage, cosmid, viral (such as adenoviral or retroviral; Mulligan, 1993, Science 260 926-932; Miller, 1992, Nature 357 455-460; Salmons et al., 1993, Hum. Gen. Ther. 4 129-141, which articles are incorporated herein by reference) or other vectors, or other agents of modification such as liposomes (Zhu et al., 1993, Science 261 209-212, which is incorporated herein by reference), viral capsids or nanoparticles (Bertling et al., 1991, Biotech. Appl. Biochem. 13 390-405, which is incorporated herein by reference), or any other mediator of modification. The use of cells as a delivery vehicle for genes or gene products has been described by Barr et al., 1991, Science 254 1507-1512 and by Dhawan et al., 1991, Science 254 1509-1512, which articles are incorporated herein by reference. Treated cells may be delivered in combination with any nutrient, growth

factor, matrix or other agent that will promote their survival in the treated subject.

[0072] Advantageously, the translational efficiencies of different codons may be determined by comparing expression of the reporter protein in a given cell or tissue type or between different cell or tissue types. One of ordinary skill in the art will thereby be able to determine a "codon preference table" for one or more cells or tissues. Comparison of codon preference tables relating to different cell or tissue types may be used to identify codons for tailoring a synthetic polynucleotide to target expression of a protein to a particular cell or tissue, as described hereinafter. Comparison of codons within a codon preference table for a particular cell or tissue type can be used to identify codons for tailoring a synthetic polynucleotide to express a protein at higher or lower levels in that cell or tissue type than a parent polynucleotide, as described hereinafter.

[0073] The invention further contemplates cells or tissues containing therein the synthetic construct of the invention, or alternatively, cells or tissues produced from the method of the invention.

3. Synthetic polynucleotides for targeting protein expression to a particular cell or tissue

[0074] The invention also provides an improved method of constructing a synthetic polynucleotide from which a protein is selectively expressible in a target cell of an organism, relative to another cell of the organism. This method is based in part on the method disclosed in copending International application PCT/AU98/00530 (the entire contents of which are hereby incorporated by reference) in which a first codon of a parent polynucleotide is replaced with a synonymous codon which has a higher translational efficiency in said target cell than

in said other cell. The improved method of the invention is characterized by selecting the first and synonymous codons by comparing translational efficiencies of individual codons in said target cell relative to said other cell using the method broadly described is Section 2.

3.1 Selection of synonymous and first codons

[0075] The present method preferably includes the step of selecting the codons such that the synonymous codon has a higher translational efficiency in said target cell or tissue ("cell or tissue" is sometimes referred to herein as "cell/tissue") relative to said one or more other cells or tissues.

[0076] A method for determining translational efficiencies of different codons in and between different cells or tissues is described in detail in Section 2. The translational efficiencies so determined can be used to identify which isocoding triplets are differentially translated between the different cells or tissues. In a typical scenario, there will be: (A) codons with higher translational efficiencies in a target cell/tissue relative to one or more other cells/tissues; (B) codons with higher translational efficiencies in the one or more other cells/tissues relative to the target cell/tissue; and (C) codons with about the same translational efficiencies in the target cell/tissue relative to the one or more other cells/tissues. Synonymous codons are selected such that they correspond to (A) codons. Preferably, a synonymous codon is selected such that it has the largest difference in translational efficiency in the target cell or tissue relative to the existing codon (sometimes referred to as a "first codon") that it replaces. Existing codons in a parent polynucleotide are preferably selected such that they do not have the same translational bias as the synonymous codons with respect to the target cell/tissue and the one or more other cell/tissue (*i.e.*, existing codons should preferably not correspond to (A) codons).

However, existing codons can have similar translational efficiencies in each of the target cell/tissue and the one or more other cells/tissues (i.e., existing codons can correspond to (C) codons. They can also have a translational bias opposite to that of the synonymous codons (i.e., existing codons can, and preferably do, correspond to (B) codons).

[0077] Suitably, a synonymous codon has a translational efficiency in the target cell/tissue that is at least 110%, preferably at least 200%, more preferably at least 500%, and still more preferably at least 1000%, of that in the other cell(s)/tissue(s). In the case of two or more synonymous codons having similar translational efficiencies in the target cell/tissue relative to the other cell(s)/tissue(s), it will be appreciated that any one of these codons can be used to replace the existing codon.

[0078] It is preferable but not necessary to replace all the existing codons of the parent polynucleotide with synonymous codons having higher translational efficiencies in the target cell/tissue compared to the other cells/tissues. Increased expression can be accomplished even with partial replacement. Suitably, the replacement step affects 5%, 10%, 15%, 20%, 25%, 30%, more preferably 35%, 40%, 50%, 60%, 70% or more of the existing codons of the parent polynucleotide.

[0079] The difference in level of protein expressed in the target cell/tissue from a synthetic polynucleotide relative to that expressed in the other cell(s)/tissue(s) depends on the percentage of existing codons replaced by synonymous codons, and the difference in translational efficiencies of the synonymous codons in the target cell/tissue relative to the other cell(s)/tissue(s). Put another way, the fewer such replacements, and/or the smaller the difference in translational efficiencies of the synonymous between the different cells/tissues, the smaller the difference in protein expression

between the target cell/tissue and the other cell(s)/tissue(s) will be. Conversely, the more such replacements, and/or the greater the difference in translational efficiencies of the synonymous codons between the different cells/tissues, the greater the difference in protein expression between the target cell/tissue and the other cell(s)/tissue(s) will be. The inventors have found in this respect that a protein can be expressed from a synthetic polynucleotide in a target cell/tissue at levels greater than 10,000-fold over those expressed in another cell/tissue.

[0080] In a preferred embodiment, the synonymous codon is a codon which has a higher translational efficiency in the target cell or tissue relative to a precursor cell or tissue of the target cell or tissue.

[0081] In an alternate embodiment, the synonymous codon is a codon which has a higher translational efficiency in the target cell or tissue relative to a cell or tissue derived from said target cell or tissue.

[0082] The two codons can be selected by measuring translational efficiencies of different codons in the target cell or tissue relative to the one or more other cells or tissues and identifying the at least one existing codon and the synonymous codon based on this measurement.

[0083] Suitably, the synonymous codon corresponds to a reporter construct from which the reporter protein is expressed in said target cell at a level that is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from the said reporter construct in said other cell.

3.2 Construction of synthetic polynucleotides

[0084] The step of replacing a synonymous codon for said first codon in a parent polynucleotide may be effected by any

suitable technique. For example, in vitro mutagenesis methods may be employed as for example discussed in Section 2.

[0085] It is not necessary to replace all the first codons of the parent polynucleotide with synonymous codons each corresponding to a codon that has a higher translational efficiency in the target cell relative to said other cell. Increased expression may be accomplished even with partial replacement. Preferably, the replacing step affects 5%, 10%, 15%, 20%, 25%, 30%, more preferably 35%, 40%, 50%, 60%, 70% or more of the existing codons of the parent nucleic acid sequence.

[0086] The parent polynucleotide is preferably a natural gene.

[0087] The parent polynucleotide may be obtained from a plant or an animal. Alternatively, the parent polynucleotide may be obtained from any other eukaryotic organism or a prokaryotic organism. In a preferred embodiment, the parent polynucleotide is obtained from a pathogenic organism. In such a case, a natural host of the pathogenic organism is preferably a plant or animal. For example, the pathogenic organism may be a yeast, bacterium or virus. However, it will be understood that the parent polynucleotide need not be obtained from the organism in which a protein is to be expressed but may be obtained from any suitable source such as from another eukaryotic or prokaryotic organism.

[0088] Suitable proteins which may be used for selective expression in accordance with the invention include, but are not limited to the cystic fibrosis transmembrane conductance regulator (CFTR) protein, and adenosine deaminase (ADA). In the case of CFTR, a parent nucleic acid sequence encoding the CFTR protein which may be utilized to produce the synthetic nucleic acid sequence is described, for example, in Riordan et al (1989, Science **245** 1066-1073), and in the GenBank database under

Accession No. HUMCFTRM, which are incorporated herein by reference.

[0089] Regulatory polynucleotides which may be utilized to regulate expression of the synthetic polynucleotide include, but are not limited to, a promoter, an enhancer, and a transcriptional terminator. Such regulatory polynucleotides are known to those of skill in the art. The construct preferably comprises at least one promoter. Suitable promoters that can be utilized to induce expression of the synthetic polynucleotides of the invention include constitutive promoters and inducible promoters.

[0090] Synthetic polynucleotides according to the invention may be operably linked to one or more regulatory sequences in the form of an expression vector.

[0091] The invention also contemplates synthetic polynucleotides encoding one or more desired portions of the protein to be expressed. A polynucleotide encodes a domain of the protein having a function associated therewith, or which is otherwise detectable, and preferably encodes at least 10, 20, 50, 100, 150, or 500 contiguous amino acid residues of the protein.

[0092] 4. Synthetic polynucleotides for enhanced protein expression in a particular cell or tissue

[0093] In contrast to differential protein expression between different cells/tissues, it will be appreciated that a synthetic polynucleotide may be tailored with synonymous codons such that expression of a protein in a target cell is enhanced. In this regard, the difference in level of protein expressed in the target cell/tissue from a synthetic polynucleotide relative to that expressed from a parent polynucleotide depends on the percentage of existing codons replaced by synonymous codons, and the difference in translational efficiencies between the

existing codons and the synonymous codons in the target cell/tissue. Put another way, the fewer such replacements, and/or the smaller the difference in translational efficiencies between the synonymous and existing codons, the smaller the difference in protein expression between the synthetic polynucleotide and parent polynucleotide will be. Conversely, the more such replacements, and/or the greater the difference in translational efficiencies between the synonymous and existing codons, the greater the difference in protein expression between the synthetic polynucleotide and parent polynucleotide will be. The inventors have found in this respect that a protein can be expressed from a synthetic polynucleotide in a target cell/tissue at levels greater than 10,000-fold than from a parent polynucleotide.

[0094] Preferably, the at least one existing codon and the synonymous codon are selected such that said protein is expressed from said synthetic polynucleotide in said target cell or tissue at a level which is at least 110%, preferably at least 200%, more preferably at least 500%, and most preferably at least 1000%, of that expressed from said parent polynucleotide in said target cell or tissue.

[0095] The invention is further described with reference to the following non-limiting examples.

EXAMPLE 1

Construction of expression vectors for determining relative codon preferences in mammalian cells.

[0100] Synthetic gfp genes were constructed in which a single artificial start codon (ATG) followed by a stretch of five identical codons is fused in frame immediately upstream of a gfp coding sequence. A reverse oligonucleotide primer (SEQ ID NO:185; sequence complementary to the termination codon *for GFP*,

is underlined), and a suite of forward oligonucleotide primers (SEQ ID NO: 126 through 184; the first codon of GFP, is underlined) were synthesized and used for PCR amplification of a humanized gfp gene (SEQ ID NO:124) (GIBCO) as template with Taq DNA polymerase (Amplification parameters: 95°C/30 sec; 52°C/30 sec; 72°C/1 min; 30 cycles). The amplified fragments have nucleic acid sequences and deduced amino acid sequences as shown in SEQ ID NO:1 through 124.

[0101] In summary, the synthetic fragments contain an artificial start codon followed by a tandem repeat of five identical codons specific for a given iso-tRNA species. The tandem repeat immediately precedes the second codon of the gfp gene. The synthetic fragments by SEQ ID NO, and encoded tandem repeat, are presented in the TABLE 1.

[0102] The amplified fragments were cloned between the EcoRI and KpnI sites of the mammalian expression vector pCDNA3 containing SV40 ori (Invitrogen) and the CMV promoter.

Transfection of COS-1 cells

[0103] COS-1 cells were grown continuously in DMEM media supplemented with 10% fetal calf serum (FCS), glutamine, penicillin and streptomycin. Cells were passaged from a 150 cm^2 flask into multiple 25 cm^2 flasks. Cells were transfected using a QIAGEN Effectene™ transfection kit (and the manufacturer's instructions, incorporated herein by reference) when confluency of the cells was between 60-80%. Briefly, 1 µg of plasmid DNA was diluted into 10 µL of filtered TE buffer and 140 µL of QIAGEN™ Buffer EC. Eight microliters of QIAGEN™ Enhancer was added followed by vortexing and incubation at room temperature for 2-5 min. QIAGEN™ Effectene (10 µL) was added followed by vortexing for 10 seconds and a further incubation at room temperature for 10 min. The cells were washed once in 1x PBS followed by re-suspension in fresh media (1 mL). *After 48 hrs,*

cells were harvested and washed in 1x PBA (phosphate buffered saline plus azide). Cells adhering to the flask were removed by scraping with a cell scraper. Cells were then filtered through a 70 μm filter before addition of 300 μL of 2% paraformaldehyde and 300 μL of 10x FCS. Cells were kept on ice in the dark until FACS analysis.

[0104] Synthetic *gfp* mRNA expression of transfected cells was tested by reverse transcriptase PCR. GFP protein expression was analyzed by confocal microscopy and flow cytometry.

Confocal microscopy

[0105] Transfected COS-1 cells were examined using a Bio-Rad MRC-600 laser-scanning confocal microscope equipped with a krypton-argon laser and filter sets suitable for the detection of fluorescein and Texas red dyes (Bio-Rad K1yK2), and a Nikon 603 PlanApo™ numerical aperture 1.2 water-immersion objective. Dual-channel confocal images and video montages of the transfected cells can be suitably composed using ADOBE PhotoShop™.

Flow cytometry

[0106] Transfected COS-1 cells were analyzed with a Becton Dickinson™ Flow cytometer Elite II. Omega Filters™ allowed detection of green fluorescence emission (EMI510/20 - collects light from 490-530 nm) and yellow fluorescence emission (EM2 550/30 - collects light form 525-580 nm) from the transfected cells.

Results

[0107] A series of 64 reporter constructs (see TABLE 1) was made and validated, in which the gfp gene is preceded in frame by a tandem repeat of 5 identical codons. Together, the series covers the entire set of isoaccepting codon triplets.

[0108] The series was transfected into a single cell line, and expression levels measured by flow cytometry (see TABLE 2). Overall, the expression level of the reporter gene constructs in the cell line varied over a range of 20-fold, according to the triplet used in the reporter construct. Repeated determinations on the same construct showed excellent inter-assay reproducibility ($r^2 = 0.9$). Variation in expression levels across the isoaccepting codons for a single amino acid ranged from 1.4-fold for valine to 13-fold for threonine, with a median of about 4-fold. Variation in expression between amino acids was of the same order of magnitude. The order of magnitude of the effect is defined as an average of 4 fold per amino acid if 5 copies are incorporated, compatible with an extreme in range of expression levels of up to $(1.6)^{200} = 10^{86}$ over an average 200-amino acid residues protein. This figure is derived as:

$$[1 + ((4-1)(\text{ range of reporter construct expression})/ 5 \text{ (no of triplets in the reporter construct)})]^{200\text{(no of amino acid residues in the protein)}}$$

and is more than sufficient to explain the observed differences in expression of mammalian genes according to codon usage.

[0100] The results presented in TABLE 2 also show that various codons in the undifferentiated epithelial cells (COS-1) have translational efficiencies at least two-fold higher or two-fold lower relative to those of their corresponding synonymous codons. Representative codons having at least a two-fold higher translational efficiency relative to at least one of their corresponding synonymous codons include aga (Arg), cgg (Arg), tgc (Cys), gga (Gly), ggc (Gly), ccg (Pro), cga (Pro), aca (Thr), acg (Thr), and act (Thr). Thus, these codons appear to be preferred for translation in the undifferentiated epithelial cells. By contrast, representative codons having at least a two-fold lower translational efficiency relative to at least one of their corresponding synonymous codons include *agg (Arg)*, tgt

(Cys), ggg (Gly), ggt (Gly), ccc (Pro), cct (Pro), and acc (Thr). These latter codons would therefore appear to be less preferred for translation in the undifferentiated epithelial cells. Accordingly, if higher protein expression is required within undifferentiated epithelial cells such as COS-1 cells, the preferred codons should be used to replace any existing codons of a parent polynucleotide encoding the protein that correspond to the less preferred codons. In this respect, a codon substitution algorithm for increasing protein expression in non-differentiated epithelial cells is presented in TABLE 3. However, if lower protein expression is required in non-differentiated epithelial cells, the less preferred codons should be used to replace any existing codons of the parent polynucleotide that correspond to the preferred codons.

[0101] The disclosure of every patent, patent application, and publication cited herein is hereby incorporated by reference in its entirety.

[0102] The present invention has been described in terms of particular embodiments found or proposed by the present inventors to comprise preferred modes for the practice of the invention. Those of skill in the art will appreciate that, in light of the present disclosure, numerous modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the invention. All such modifications are intended to be included within the scope of the appended claims.

TABLE 1

[0103] Synthetic gfp constructs are tabulated by SEQ ID NO and by the codon corresponding to the tandem repeat of five identical codons immediately upstream of the gfp gene.

TABLE 2

[0104] Mean fluorescence intensities of up to four different samples of transiently transfected COS-1 cells are shown (Green mean 1-4). Synthetic gfp constructs are tabulated by SEQ ID NO and by the codon corresponding to the tandem repeat immediately upstream of the gfp gene.

TABLE 3

[0105] Input codons and output codons represent, respectively, synonymous codons and existing (i.e., "first") codons according to the invention. Change means an actual change of a codon.

TABLES

TABLE 1

Synthetic fragments and tandem repeats encoded thereby.

SEQ ID NO	Tandem repeat	SEQ ID NO	Tandem repeat
1	Ala (GCA) x 5	65	Leu (CTT) x 5
3	Ala (GCC) x 5	67	Leu (TTA) x 5
5	Ala (GCG) x 5	69	Leu (TTG) x 5
7	Ala (GCT) x 5	71	Lys (AAA) x 5
9	Arg (AGA) x 5	73	Lys (AAG) x 5
11	Arg (AGG) x 5	75	Phe (TTT) x 5
13	Arg (CGA) x 5	77	Phe (TTC) x 5
15	Arg (CGC) x 5	79	Pro (CCC) x 5
17	Arg (CGG) x 5	81	Pro (CCG) x 5
19	Arg (CGT) x 5	83	Pro (CCT) x 5
21	Asn (AAC) x 5	85	Pro (CGA) x 5
23	Asn (AAT) x 5	87	Ser (AGC) x 5
25	Asp (GAC) x 5	89	Ser (AGT) x 5
27	Asp (GAT) x 5	91	Ser (TCA) x 5
29	Cys (TGC) x 5	93	Ser (TCC) x 5
31	Cys (TGT) x 5	95	Ser (TCG) x 5
33	Gln (CAA) x 5	97	Ser (TCT) x 5
35	Gln (CAG) x 5	99	Thr (ACA) x 5
37	Gly (GAA) x 5	101	Thr (ACC) x 5
39	Gly (GAG) x 5	103	Thr (ACG) x 5

SEQ ID NO	Tandem repeat	SEQ ID NO	Tandem repeat
41	Gly (GGA) x 5	105	Thr (ACT) x 5
43	Gly (GGC) x 5	107	Trp (TGG) x 5
45	Gly (GGG) x 5	109	Tyr (TAT) x 5
47	Gly (GGT) x 5	111	Tyr (TAC) x 5
49	His (CAC) x 5	113	Val (GTA) x 5
51	His (CAT) x 5	115	Val (GTC) x 5
53	Ile (ATA) x 5	117	Val (GTG) x 5
55	Ile (ATC) x 5	119	Val (GTT) x 5
57	Ile (ATT) x 5	121	Stop (TAA) x 5
59	Leu (CTA) x 5	122	Stop (TAG) x 5
61	Leu (CTC) x 5	123	Stop (TGA) x 5
63	Leu (CTG) x 5	124	control

TABLE 2

GFP protein expression in transiently transfected COS-1 cells

SEQ ID NO	Codon	[DNA] (µg/mL)	Green mean 1	Green mean 2	Green mean 3	Green mean 4	Average
1	Ala (GCA)	1.07	45.70	54.40			50.05
3	Ala (GCC)	1.10	43.70	50.00			46.85
5	Ala (GCG)	0.03	28.50	42.40			35.45
7	Ala (GCT)	0.56	11.60	48.30			29.95
9	Arg (AGA)	0.90	29.00	33.00			31.00
11	Arg (AGG)	0.34	7.35	2.88			5.12
13	Arg (CGA)	1.00	18.30	14.20			16.25
15	Arg (CGC)	0.86	14.60	16.00			15.30
17	Arg (CGG)	1.00	22.50	20.60			21.55
19	Arg (CGT)	0.68	21.70	32.20			26.95
21	Asn (AAC)	0.02					
23	Asn (AAT)	0.38	28.30	8.22			18.26
25	Asp (GAC)	0.46	24.90	17.80			21.35
27	Asp (GAT)	1.39	14.50	18.90			16.70
29	Cys (TGC)	0.68	21.90	16.10			19.00
31	Cys (TGT)	1.14	5.95	5.89			5.92
33	Gln (CAA)	0.28	26.50	43.50			35.00
35	Gln (CAG)	1.98	44.70	48.60			46.65
37	Glu (GAA)	0.60	10.30	22.70			16.50
39	Glu (GAG)	0.43	3.86				
41	Gly (GGA)	0.33	28.80	36.30			32.55
43	Gly (GGC)	1.62	17.80	28.10			22.95
45	Gly (GGG)	1.15	6.43	4.96			5.70
47	Gly (GGT)	1.39	7.12	4.02			5.57
49	His (CAC)	1.62	29.90	39.70			34.80
51	His (CAT)	1.69	43.40	37.20			40.30

SEQ ID NO	Codon	[DNA] (µg/mL)	Green mean 1	Green mean 2	Green mean 3	Green mean 4	Average
53	Ile (ATA)	0.69	2.76	3.98			3.37
55	Ile (ATC)	1.52	4.12	2.83			3.48
57	Ile (ATT)	1.77	3.19	3.16			3.18
59	Leu (CTA)	0.10	15.00	3.01	5.26	2.44	6.43
61	Leu (CTC)	1.74	2.70	2.92	2.56		2.73
63	Leu (CTG)	0.41	2.80	7.51	2.63		4.31
65	Leu (CTT)	1.43	3.17	3.56	2.70		3.14
67	Leu (TTA)	0.62	3.85	3.91	2.66		3.47
69	Leu (TTG)	0.70	2.87	4.63	2.85		3.45
71	Lys (AAA)	0.10	11.90	8.24			10.07
73	Lys (AAG)	0.56	19.20	16.00			17.60
75	Phe (TTT)	2.28	2.67				
77	Phe (TTC)	1.65	4.35				
79	Pro (CCC)	0.40	12.00	8.95			10.48
81	Pro (CCG)	0.13	17.40	25.40			21.40
83	Pro (CCT)	0.40	10.60	9.89			10.25
85	Pro (CGA)	0.17	27.20	34.80			31.00
87	Ser (AGC)	0.03	62.40				
89	Ser (AGT)	0.81	23.10				
91	Ser (TCA)	0.08	30.70	37.20			33.95
93	Ser (TCC)	1.68	32.90				
95	Ser (TCG)	1.58	60.00				
97	Ser (TCT)	0.62	26.80	40.70			33.75
99	Thr (ACA)	1.70	37.80	39.90			38.85
101	Thr (ACC)	7.69	3.48	2.75			3.12
103	Thr (ACG)	1.06	36.10	44.10			40.10
105	Thr (ACT)	1.42	38.80	42.60			40.70
107	Trp (TGG)	1.19	5.21	4.29			4.75
109	Tyr (TAT)	0.02					
111	Tyr (TAC)	1.07	12.00	15.00			13.50
113	Val (GTA)	0.16	10.50	3.81			7.16

SEQ ID NO	Codon	[DNA] (µg/mL)	Green mean 1	Green mean 2	Green mean 3	Green mean 4	Average
115	Val (GTC)	0.66	15.20	4.55	3.65	5.06	7.12
117	Val (GTG)	0.10	9.17	4.29	7.03	2.36	5.71
119	Val (GTT)	0.49	14.10	2.63	3.70	2.49	5.73
121	stop (TAA)	1.88	39.40	35.30			37.35
122	stop (TAG)	2.86	2.88	3.28			3.08
123	stop (TGA)	0.02					
124 GFP alone			9.34	61.60	30.40	55.00	39.09
control			2.33	2.21	2.16	2.00	2.18

TABLE 3

Substitution algorithm used for high level expression in non-differentiated epithelial cells

Input Codon	Output Codon	Amino Acid	Change
AAA	AAG	LYS	Yes
AAC	AAC	ASN	No
AAG	AAG	LYS	No
AAT	AAC	ASN	Yes
AAU	AAC	ASN	Yes
ACA	ACC	THR	Yes
ACC	ACC	THR	No
ACG	ACC	THR	Yes
ACT	ACC	THR	Yes
ACU	ACC	THR	Yes
AGA	AGG	ARG	Yes
AGC	AGC	SER	No
AGG	AGG	ARG	No
AGT	AGC	SER	Yes
AGU	AGC	SER	Yes
ATA	ATC	ILE	Yes
ATC	ATC	ILE	No
ATG	ATG	MET	No
ATT	ATC	ILE	Yes
AUA	ATC	ILE	Yes
AUC	ATC	ILE	No
AUG	ATG	MET	No
AUU	ATC	ILE	Yes
CAA	CAG	GLN	Yes
CAC	CAC	HIS	No
CAG	CAG	GLN	No
CAT	CAC	HIS	Yes
CAU	CAC	HIS	Yes
CCA	CCC	PRO	Yes
CCC	CCC	PRO	No
CCG	CCC	PRO	Yes
CCT	CCC	PRO	Yes
CCU	CCC	PRO	Yes

Input Codon	Output Codon	Amino Acid	Change
CGA	CGC	ARG	Yes
CGC	CGC	ARG	No
CGG	CGC	ARG	Yes
CGT	CGC	ARG	Yes
CGU	CGC	ARG	Yes
CTA	CTG	LEU	Yes
CTC	CTG	LEU	Yes
CTG	CTG	LEU	No
CTT	CTG	LEU	Yes
CUA	CTG	LEU	Yes
CUC	CTG	LEU	Yes
CUG	CTG	LEU	No
CUU	CTG	LEU	Yes
GAA	GAG	GLU	Yes
GAC	GAC	ASP	No
GAG	GAG	GLU	No
GAT	GAC	ASP	Yes
GAU	GAC	ASP	Yes
GCA	GCC	ALA	Yes
GCC	GCC	ALA	No
GCG	GCC	ALA	Yes
GCT	GCC	ALA	Yes
GCU	GCC	ALA	Yes
GGA	GGC	GLY	Yes
GGC	GGC	GLY	No
GGG	GGG	GLY	No
GGT	GGC	GLY	Yes
GGU	GGC	GLY	Yes
GTA	GTG	VAL	Yes
GTC	GTG	VAL	Yes
GTG	GTG	VAL	No
GTT	GTG	VAL	Yes
GUA	GTG	VAL	Yes
GUC	GTG	VAL	Yes
GUG	GTG	VAL	No
GUU	GTG	VAL	Yes
TAA	TAA	XXX	No
TAC	TAC	TYR	No
TAG	TAG	XXX	No

Input Codon	Output Codon	Amino Acid	Change
TAT	TAC	TYR	Yes
TCA	TCC	SER	Yes
TCC	TCC	SER	No
TCG	TCC	SER	Yes
TCT	TCC	SER	Yes
TGA	TGA	XXX	No
TGC	TGC	CYS	No
TGG	TGG	TRP	No
TGT	TGT	CYS	No
TTA	CTG	LEU	Yes
TTC	TTC	PHE	No
TTG	CTG	LEU	Yes
TTT	TTC	PHE	No
UAA	TAA	XXX	No
UAC	TAC	TYR	No
UAG	TAG	XXX	No
UAU	TAC	TYR	Yes
UCA	TCC	SER	Yes
UCC	TCC	SER	No
UCG	TCC	SER	Yes
UCU	TCC	SER	Yes
UGA	TGA	XXX	No
UGC	TGC	CYS	No
UGG	TGG	TRP	No
UGU	TGT	CYS	No
UUA	CTG	LEU	Yes
UUC	TTC	PHE	No
UUG	CTG	LEU	Yes
UUU	TTC	PHE	Yes